(12)

EUROPEAN PATENT APPLICATION

(21) Application number: **84630096.0**

(22) Date of filing: **15.06.84**

(51) Int. Cl.⁴: **G 10 L 7/06**

---

(72) Inventor: **Dewhurst, David John**
37 Cullwell Avenue
Mitcham 3132 Victoria(AU)

(72) Inventor: **Ng, Chee Wei**
10 JN, Union 3
Sentul Kuala Lumpur 13-06(MY)

(72) Inventor: **Hughes, Murray Allan**
25 Gwynn Street
Mount Waverley 3149 Victoria(AU)

(72) Inventor: **Johnson, Donald Archibald Harley**
Lizbeth Avenue
Mitcham 3132 Victoria(AU)

(74) Representative: **Waxweiler, Jean et al,**
OFFICE DENNEMEYER S.à.r.l. 21-25 Allée Scheffer
P.O.Box 41
L-2010 Luxembourg(LU)

---

(54) **Signal processing.**

(57) The disclosed system for extracting desired information from a speech signal includes means for taking overlapping samples of an utterance, computer means programmed to test each sample to determine whether it is voiced or unvoiced and for performing the following operations on each voiced sample:

applying a 30 ms. Hamming window to smooth the edge of the signal and to ensure that false artifacts will not be present in the following processing stage, obtaining a magnitude spectrum using at least 1024 points Fast Fourier transform, obtaining the log of the magnitude spectrum, compressing the spectrum, performing a three-point filter algorithm a suitable number of times, expanding the spectrum so obtained and locating the dominant peaks in the resulting spectrum to give the information content contained in said speech signal. The specification also discloses the time equivalent of the above method.
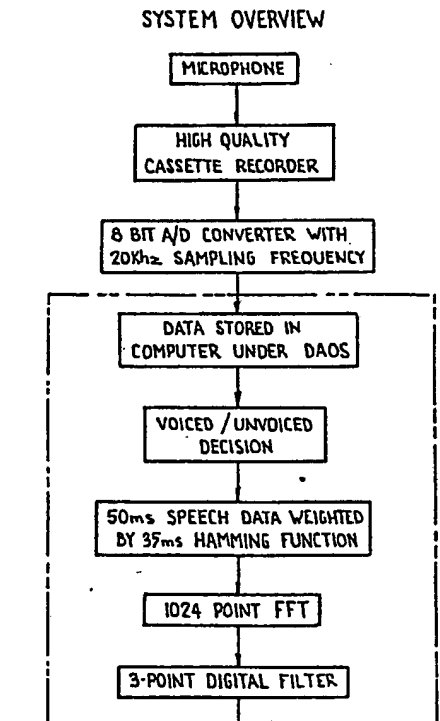
EP 0 132 216 A

./...

SYSTEM OVERVIEW

MICROPHONE

HIGH QUALITY
CASSETTE RECORDER

8 BIT A/D CONVERTER WITH
20Khz SAMPLING FREQUENCY

DATA STORED IN
COMPUTER UNDER DAOS

VOICED / UNVOICED
DECISION

50ms SPEECH DATA WEIGHTED
BY 35ms HAMMING FUNCTION

1024 POINT FFT

3-POINT DIGITAL FILTER

FIG. 7A.

PEAK PICKING ALGORITHM

AMPLITUDE AND FREQUENCY OF
FOUR SPECTRAL PEAKS STORED

NEXT PITCH PERIOD
IS THEN ANALYSED
ASYNCHRONOUSLY

FOUR SPECTRAL POINTS PLACE
ON FREQUENCY DOMAIN AND
3-PT DIGITAL FILTERING PERFORM

MULTIPLY BY RANDOM
NOISE IF UNVOICED

INVERSE FFT TO
GENERATE WAVEFORM
FOR EACH PITCH PERIOD

GENERAL
PURPOSE COMPUTER

HIGHLY INTELLIGIBLE
SPEECH REPRODUCTION
AT 8KBIT/SEC D/A CONVERTER

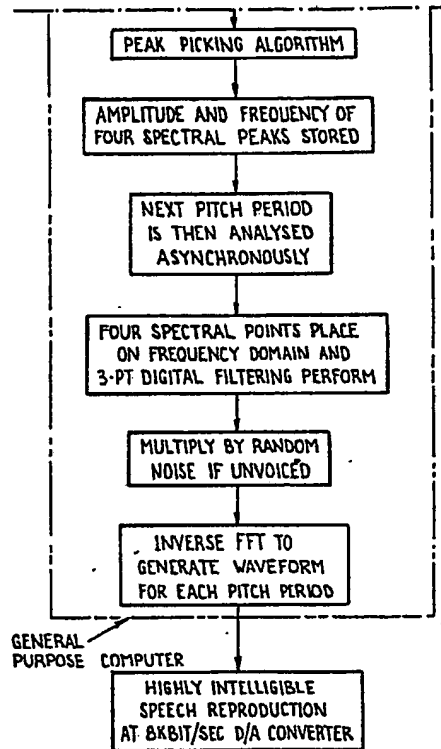. FIG. 7B .

TITLE: SIGNAL PROCESSING.

BACKGROUND OF THE INVENTION

This invention relates to a system for the processing of signals to extract desired information. The invention is particularly applicable to the extraction of the desired information content from a
5   received speech signal for subsequent use in activating or stimulating an implantable hearing prosthesis or for other purposes.

The variability of speech signals between speakers of the same utterance (as shown in Figure 1)
10   has been a major problem faced by all speech scientists. However, the fact that human auditory system is capable of extracting relevant speech information from widely varying speech signals has baffled speech researchers for decades. The
15   information must of course be present in the signal but thus far researchers in this field have been unable to devise a system for reliably extracting the information from a speech signal.

The retrieval of text from voice involving
20   recognition of unrestricted speech is still considered to be far beyond the current state of the art. What is being attempted is automatic recognition of words from restricted speech. Even so, the reliability of these ASR (Automatic Speech Recognition) systems is
25   unpredictable. One report ("Selected Military Applications of ASR Technology" by Woodward J.P. & Cupper E.J., IEEE Communications Magazine, 21, 9

December 1983, pg 35-41) lists eighty different
factors which can affect their reliability. Such
advances in ASR as have been achieved have arisen
more from improved electronics and microprocessor
5   chips than from the development of any new technology
for ASR.

In considering this question, the present
inventors have given consideration to the manner in
which the auditory system handles widely varying
10   speech signals and extracts the information required
to make the speech signal intelligible. .When sounds
of speech are transmitted to the higher centres of
the brain by means of the auditory system it under-
goes several physiological processes.

15

When speech signals arrive at the middle
ear, a mechanical gain control mechanism acts as an
automatic gain control function to limit the dynamic
range of the signal being analysed. According to
the temporal-place representation, the discharge
20   patterns of auditory nerve-fibres show stronger
phase locking behaviour to spectral peaks than
locking to other harmonics of the stimulus. At
physiological sound level, synchrony to dominant
spectral peaks saturates and responses to pitch
25   harmonics are suppressed. The resulting effect is
such that the rough appearance of the pitch harmonics
are masked out.

SUMMARY OF THE INVENTION

The present inventors therefore determined
30   that if the pitch information (that is, the speaker
attributes, which contain no real information) such as
pitch frequency, harmonics components thereof and other
minor speaker attributes, could be removed from the
speech signal, the remaining signal would contain the
35   information necessary for understanding the utterance
contained in the complex speech signal thereby

resulting in a signal usable to stimulate a hearing
prosthesis or for other purposes, such as, speech
recognition by computer, speech synthesis, and speech
bandwidth compression for rapid transmission of

5   speech.

In its broadest aspect therefor, the
invention provides a system for extracting desired
information from a speech signal including means for
performing the essential steps of removing from or

10 suppressing in the speech signal at least the sig-
nificant components relating to pitch frequency, and
identifying and tracking in time the spectral peaks
of the resulting signal.

BRIEF DESCRIPTION OF THE DRAWINGS

15   In the drawings:

Figure 1 is a plot showing the variability
in the speech signals of two
different speakers of the same
utterance;

20   Figures 2 and 3 are spectral plots of
frequency against time of the
utterance shown in Figure 1 again
showing the variability of the
signals;

25   Figure 4 shows the effect of applying a
smoothing algorithm to the signal;

Figures 5 and 6 show plots of the spectral
peaks produced by the smoothing
shown in Figure 4;

30   Figure 7 is a schematic representation of
one signal processing method
embodying the invention;

Figures 8 to 15 show the steps in the
processing method as applied to a

35   specific utterance;

Figure 15A is a three-dimensional plot of
the spectral peak variation
against time of the utterance
'Boat';

5   Figure 16 is a schematic representation of
a real-time speech synthesizer;

Figures 17 and 17A show typical line
representations of the utterance
'Melbourne' to be used in the
10   synthesizer described in
Figure 16; and

Figures 18 to 23 show the steps in an
alternative processing method.

DESCRIPTION OF PREFERRED EMBODIMENTS

15   The aim of many techniques of analysing
speech signals is to characterize the temporal
variation of the amplitude spectra of short intervals
of a word. A digital method of producing a frequency
spectrum of a short time interval by means of the
20   Fast Fourier Transform (FFT) will yield a "messy"
spectrum caused by pitch harmonics. Plots of these
spectral variations against time shown in Figures 2
and 3 will be seen to be masked by the dominant pitch
harmonics.

25   A smoothing algorithm is performed on the
signal "noise" in the spectrum and is filtered and the
centre frequency and amplitude of the four locally
dominant spectral peaks are able to be picked out
(see Figure 4). Plots of these spectral peaks against
30   time are shown in Figures 5 and 6. The similarities
in these plots between speakers are clearly evident
particularly in the direction of movement of the
spectral peak tracks. Unlike formants, these spectral
lines are discontinuous and their movements cover a
35   wider bandwidth. There is little doubt that this

concept of processing is the first step towards speech perception.

Using the information acquired by the above process, a reverse processing technique can be used to resynthesize highly intelligible speech on a digital computer. The same information can be displayed in two dimensions as line patterns and by means of an optical reader these lines may be converted back into speech frequencies. Using this concept it can be demonstrated that intelligible speech can be produced on a real-time hardware synthesizer even without amplitude variations.

It is envisaged that this method of speech processing can offer data rate reduction of the order of 1:40 without subjectively losing much fidelity in speech transmission.

Various methods of achieving the above described ends may be applied to the speech signal and two different approaches will now be described in greater detail.

In the first processing approach, the signal is received and processed in the manner schematically outlined in Figure 7. The process begins with the sampling of a prefiltered speech signal at a rate of about 20000 samples per second. The sampled speech is then analyzed in segments of duration 50ms. Successive 50ms segments are analyzed at 10ms intervals so that there is an overlap of adjacent segments to provide the necessary continuity. The processing technique may be better understood by considering the following example of an actual speech signal conveying the word 'boat'. The process involves the following steps of:

(a) Taking a 50ms speech sample from the word BOAT("O"), (Fig. 8)

(b) Applying the voiced/unvoiced test (as described further below),

(c) Applying a 30ms Hamming window (Fig. 9) to smooth the edge of the signal and to ensure that false artifacts will not be present in the following processing stage,

(d) Obtaining a magnitude spectrum using at least 1024 points Fast Fourier Transform (Fig. 10),

(e) Log of the magnitude spectrum (Fig. 11),

(f) Spectrum compression (Fig. 12),

(g) Three-point filter algorithm is applied a suitable number of times, (Fig. 13),

(h) Spectrum is expanded as in (Fig. 14),

(i) Four dominant peaks are located as described in the mathematical details given below (Fig. 15).

Figure 15A shows the spectral peaks extracted by the above method in a three-dimensional plot.

When a 50ms segment is transformed by the discrete Fourier Transform process, the resulting spectrum consists of a number of spectral lines occurring at frequencies which are multiples of 20Hz. The distribution of amplitudes of these lines across the frequency range, however, indicates the true distribution of spectral energy of the speech segment. The human observer can pick out the peaks of the spectral energy (i.e. the positions where the energy distribution has obvious maxima) by eye with little difficulty (see Figs. 2 and 3). The above described technique enables a computer to perform a similar task but the process is quite involved especially as care has to be taken to eliminate artifacts of the sampling process which have nothing to do with the original speech segment. The process also smooths out other features of the

spectrum dependant on pitch pulse spectral energy, speaker specific characteristics and the like.

The discrete Fourier Transform is performed by the Fast Fourier Transform routine.

$$X_n(e^{j2\pi/N}) = \sum_{m=0}^{N-1} y(n-m)x(m)e^{-j2\pi m/N}$$

5

N=1024 points

y(n) is a suitable raised cosine window.

The three point filter algorithm is represented by:

$$p(n) = \left[\frac{p(n-1)}{4} + \frac{p(n)}{2} + \frac{p(n+1)}{4} \right]$$

10   For a function as shown below

$$X(k) = \left[x(k-1)/4 + x(k)/2 + x(k+1)/4 \right]$$

the corresponding time sequence would be

$$= W_N^n \, x(n)/4 + W_N^{-n} \, x(n)/4 + x(n)/2 \quad \text{where } W_N = e^{-j(2\pi/N)}$$

$$= e^{-2\pi jn/N} x(n)/4 + e^{2\pi jn/N} x(n)/4 + x(n)/2$$

15   $$= 1/2\, x(n) \left[1 + \cos 2\pi n/N \right]$$

i.e.

$$X(k) <> 1/2 x(n)\left[1+\cos(2\pi n/N)\right] = F_1\left[X(k)\right]$$

$$F_m(X(k)) = x(n)\left[1 + \cos(2\pi n/N)\right]^m$$

Thus the time domain equivalent of a three-
20   point filtering on the frequency domain is multiplication by

$$x(n)\left[1 + \cos(2\pi n/N)\right]$$

Frequency compression on the magnitude spectrum is represented by:

25   $$p(n) = p(3n) \quad \text{where } n = 1 \text{ to } 350$$

1024 points are compressed to 350 points by sampling

every third point.

The second derivative peak picking algorithm is
represented by:

$$dy/dx = [p(n)-p(n-1)] \quad = 0 = p'(n)$$

5

$$p''(n) = [p'(n)-p'(n-1)] = negative$$

When both these conditions are met the
location of the peak is noted. A maximum of seven
peaks can be located in the spectrum but only the
four largest are selected.

10   A speech signal may be regarded as

$$VOICED \quad when \quad Ls = 1/M \sum_{N=N}^{N+M-1} a(n) \quad is \ large$$

$$and \ as \ UNVOICED \ when \quad Ls = 1/M \sum_{n=M}^{N+M-1} a(n) \quad is \ small$$

$$AND \quad Ld = 1/M \sum_{n=N}^{N+M-1} a(n+1)-a(n) \ is \ significant$$

where Ls = absolute average level of 30ms of speech
15   Ld = absolute average level of 30ms of the
          differenced signal.

A voiced/unvoiced decision is made depending
on the nature of the source of excitation of sounds.
A voiced sound is perceived when the glottis is vibrating
20   at a certain pitch causing pulses of air to excite the
natural resonating cavities of the vocal tract.
Unvoiced sounds are produced by a turbulent flow of air
caused by a constriction at some point in the vocal
tract. In analysing speech a decision is required to
25   distinguish these so that a correct source of excitation
can be used during synthesis. An algorithm can be
written to define a voiced speech when the absolute

average signal is high and unvoiced when it is rapidly
varying and of a small amplitude. If a signal sample
is determined to be unvoiced it is disregarded in the
analysis process.

5   The method employed limits the spectral
peak resolution of the resulting spectrum. However,
it is found that the centre frequency and the amplitude
of four locally dominant spectral peaks are sufficient
information for the auditory system to characterise
10   the short term acoustic properties that distinguish
one speech sound from another.

        It is also known that auditory neural
activities adapt themselves (neural adaption) whereby
a high intensity stimulus will quickly reach
15   saturation level. A similar process of adaptive
frequency equalization is done on the frequency spectrum
by transforming it to a log scale to ensure that the
more important higher frequency components are not
lost while keeping the stronger low frequency components
20   within dynamic range. Furthermore, only the magnitude
spectrum need be considered, since the cochlea is
unable to resolve signal phase components.

        A property of the cochlear and neural system
is that it can only respond to changes of a time constant
25   of the order of 10 ms. It is thus necessary that the
processing technique employed extracts and updates its
information rate every 10 ms.

        Using the above method of processing, the
information extracted, that is, the time variation of
30   the spectral peaks movements, can be used as inputs to
an implantable hearing prosthesis (such as described
in Australian specifications AU-A 41061/78 and
AU-A 59812/80 to mimic the function of the cochlea.

        The same information can be used for
35   speech recognition as illustrated in spectral plots

against time. Thirdly, using the information acquired,
a reverse processing technique can resynthesize
intelligible speech either on a digital computer or
on a real-time hardware synthesizer.

5   During resynthesis, each spectral peak
position is relocated in the frequency domain, without
regard to phase. Three-point digital smoothing is
done on these points to spread the spectrum. This
would produce a decaying waveform for every pitch
10   period generated in the time domain. The inverse FFT
is performed and a data length corresponding to a
pitch period is extracted.

          For unvoiced speech, the spectrum is
multiplied by a random phase function prior to inverse
15   FFT. A 600Hz bandwidth for the noise spectral peak
is satisfactory. The next set of data is decoded
similarly until the end of the utterance.

          In designing a real-time speech synthesizer
as shown schematically in Figure 16, one must consider
20   a method of converting these spectral lines into sine
waves of frequencies from 0.3Khz. A linear 256-pixels
RETICON chip is used. It is enclosed inside a
commercial camera with focus and aperture size adjust-
ments. The camera is mounted on an optical bench with
25   a rotating drum at right-angles to it. Four controlled
oscillators using X2206 function generator chips are
required.

          A start pulse every 10ms is used to start the
count to locate the position of each line. A maximum of
30   four lines may be identified, and the position of each
line is decoded as an 8-bit address. The address is
then latched, so that the D/A of each line is in
continuous operation throughout the 10ms period. If the
position of the line changes in the next 10ms, a 'new'
35   address is latched. If the line disappears an analogue

switch will disable the oscillator.

The D/A comprises a ladder-network to allow up to 8-bits of accuracy in determining the current flow into the X2206 oscillator chip.

5          $Frequency = 320\ I(mA)/C(\mu F)\quad Hz$

Having set a fixed capacitance, the frequency generated by the chip is only dependent of the position of the line. The output from the four oscillators are summed and multiplied by a

10 triangular wave function with an offset. This procedure will generate a pitch period as well as spreading the spectrum wider as it appears in normal speech.

A typical line input representing the

15 word 'Melbourne' is shown in Figure 17. In Figure 17A the base line has been removed since this does not contain any information and may be replaced by a straight line as shown. It has been established above that the variation with time of the frequencies

20 at which the spectral energy maxima occur contains all the information necessary to resynthesize the spoken words. Moreover we have found that the changes in amplitude of the maxima are unimportant in resynthe- sizing understandable words (though they may be

25 important in speaker identification) and the actual pitch frequency used is not critical at all. In this respect in particular the approach of this invention differs from that of others which endeavour to determine pitch frequency accurately. In the

30 resynthesis process the outputs of three or four tone generators whose frequencies are controlled by the frequency peak 'tracks', are combined, and finally a tone representing the pitch frequency added.in. This last step is not actually essential for intelligibility,

35 but improves realism.

An alternative processing method, which can be shown to be mathematically "the time" equivalent of the above method, will now be briefly explained with reference to Figures 18 to 23. This processing
5   method involves the following steps:

(a) A sample of the time waveform of the same utterance BOAT. (Fig. 18),

(b) Time expansion of the speech sample (Fig. 19),

(c) Applying a window of the form $(1 + \cos(\Pi t/T))^N$
10   (Fig. 20),

(d) Resulting waveform after windowing (Fig. 21),

(e) Obtaining a magnitude spectrum using at least 1024 points Fast Fourier transform,

(f) Log of the magnitude spectrum (Fig. 22), and

15 (g) Four dominant peaks are located (Fig. 23).

As in the case of the embodiment of Figure 7, each of the above operations are performed using a suitably programmed general purpose computer.

As mentioned above, other methods of achieving
20 the same results may be easily devised using standard mathematical procedures. Similarly, the processing techniques by which the above described alternative processing steps may be performed in a computer will be well understood by persons skilled
25 in the art and will not therefore be described in greater detail in this specification. The manner in which the extracted information is utilized will vary according to the application and although the processing technique was developed with application
30 to a hearing prosthesis in mind, the technique clearly has wider application, several of which have been indicated above. Other applications include:

Control of plant and machines by spoken command.
Aids for handicapped - voice operated wheel
35   chairs, voice operated word processors and braille writing systems.

Voice operation of computers.

Automatic information systems for public use
activated by spoken commands.

Automatic typescript generation from speech.

CLAIMS:

1.   A system for extracting desired information
from a speech signal, including means (Fig. 7) for performing
the essential steps of removing from or suppressing
in the speech signal at least the significant compon-
ents relating to pitch frequency, and identifying and
tracking in time the spectral peaks of .the resulting
signal.

2.   The system of claim 1, wherein said removal
or suppression step comprises the steps of taking
samples of the speech signal to be processed and
filtering the samples to remove or suppress the pitch
components therein whereby the locally dominant spectral
peaks are more readily able to be located and tracked.

3.   The system of claim 2, wherein the filtering
of said signal is performed in accordance with a
three point filter algorithm.

4.   The system of claim 2 or 3, wherein said signal
is Fourier transformed prior to said filtering.

5.   The system of claim 4, wherein each signal
component is tested to determine whether it is voiced
or unvoiced and if unvoiced, said signal component is
not subjected to Fourier transformation or filtering.

6.   The system of claim 4 or 5, wherein a Hamming
window is applied to each signal component before
Fourier transformation to smooth the edge of the
signal and to ensure that false artifacts will not
be present in the following processing stage.

7.   The system of claim 1, including means for
performing the following steps:
  (a)   taking overlapping samples of said speech.
          signal,

(b)   testing each sample to determine whether
        the sample is voiced or unvoiced, and
        performing the following steps in connection
        with each voiced sample,

(c)   applying a Hamming window to each sample,

(d)   obtaining a magnitude spectrum by per-
        forming a Fast Fourier transform on each
        sample,

(e)   obtaining the log of the magnitude spectrum
        of each sample,

(f)   compressing the spectrum so obtained,

(g)   performing a three-point filter algorithm
        on the compressed sample a plurality of
        times,

(h)   expanding the spectrum so obtained, and

(i)   locating the dominant peaks in said
        expanded spectrum.

8.   The system of claim 2, wherein said filtering
step comprises applying a low pass filtering function to
each signal sample.

9.   The system of claim 8 wherein said filtering
function is of the form $(1 + \cos (\pi t/T))^N$

10.   The system of claim 8 or 9 further including
the step of Fourier transforming said component following
filtering.

11.   The system of claim 1, including means for
performing the following steps:

(a)   overlapping samples of the time waveform of
        the speech signal are taken,

(b)   each sample is time expanded,

(c)   a filtering function of the form
        $(1 + \cos (\pi t/T))^N$ is applied to each sample,

(d)   the resulting signal is Fast Fourier
        transformed,

(e)   the log of the resulting magnitude spectrum
         is obtained and,

(g)   the dominant spectral peaks are located.

12.   A system for synthesizing intelligible speech
comprising means (Fig. 16) for storing a representation of
said spectral peak information extracted by the system according
to any preceding claim, and means (Fig. 16) for utilizing said
spectral peak information to generate a synthesized utterance.

13.   The system of claim 12 further comprising tone
oscillator means having frequencies generally corresponding
to each said spectral peak, means for varying the applied
voltage producing each tone oscillation in accordance with
the detected time variations in each spectral peak.

14.   The system of claim 13 further comprising the
addition of a tone representing pitch frequency to improve
realism in the synthesized speech.

15.   A method of extracting desired information from
a speech signal comprising the steps of removing from or
suppressing in the speech signal at least the significant
components relating to pitch frequency, and identifying and
tracking in time the spectral peaks of the resulting signal.

16.   A method of synthesizing intelligible speech
comprising the steps of storing a representation of said
spectral peak information extracted according to the method
of claim 15 and utilizing said spectral peak information
to generate a synthesized utterance.
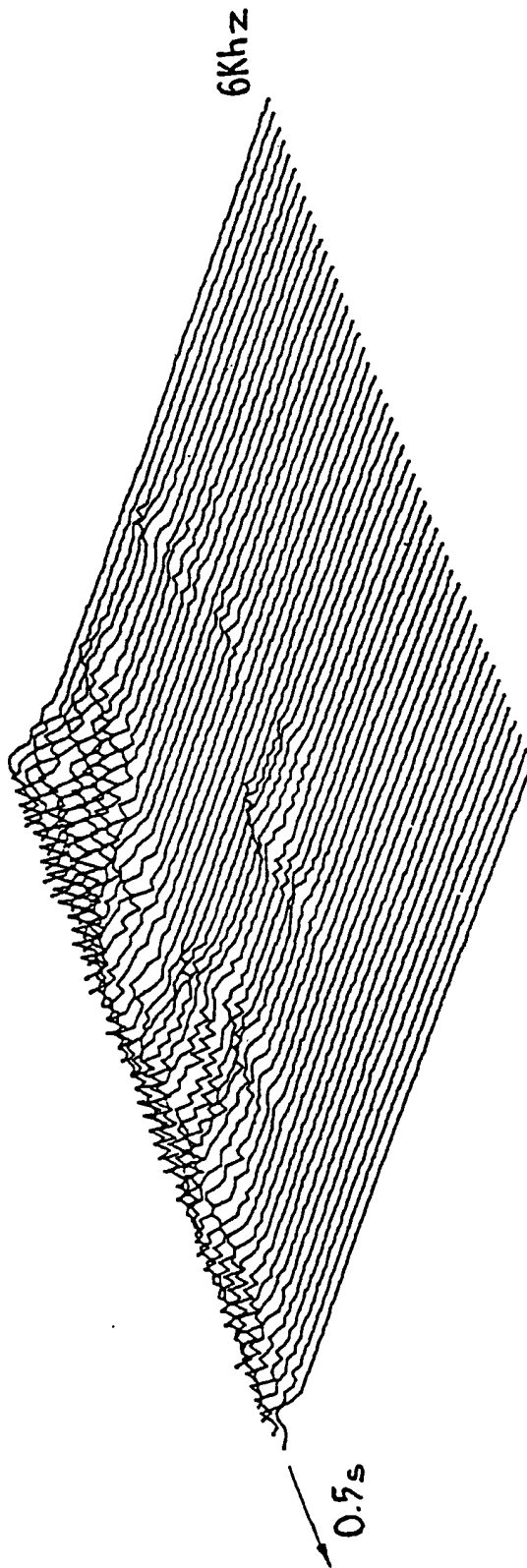
FIG. 1.

JOHN
"MELBOURNE"

0s

12.5 ms

BOB
"MELBOURNE"

0s

12.5 ms

0132216

6Khz

JOHN
"MELBOURNE"

0.5s

FIG_2_

0132216

FIG - 3 -

6Khz

0.7s

BOB
"MELBOURNE"

0db

-60db

10Khz

FIG. 4.

0db

SPECTRAL PEAK (S0)

(S1)

(S2)

(S3)

-60db

10Khz

FIG. 17.

FIG. 17A.

FIG. 5.

6Khz

0.5s

BOB
"MELBOURNE"

6/13

FIG. 6

6Khz

JOHN
"MELBOURNE"

0.5s

SYSTEM OVERVIEW

MICROPHONE

HIGH QUALITY
CASSETTE RECORDER

8 BIT A/D CONVERTER WITH
20Khz SAMPLING FREQUENCY

DATA STORED IN
COMPUTER UNDER DAOS

VOICED / UNVOICED
DECISION

50ms SPEECH DATA WEIGHTED
BY 35ms HAMMING FUNCTION

1024 POINT FFT

3-POINT DIGITAL FILTER

FIG. 7A.

```
┌─────────────────────────────────┐
│    PEAK  PICKING  ALGORITHM     │
└─────────────────────────────────┘
               │
               ▼
┌─────────────────────────────────┐
│   AMPLITUDE AND  FREQUENCY OF    │
│   FOUR SPECTRAL  PEAKS STORED    │
└─────────────────────────────────┘
               │
               ▼
┌─────────────────────────────────┐
│      NEXT PITCH PERIOD           │
│      IS THEN  ANALYSED           │
│       ASYNCHRONOUSLY             │
└─────────────────────────────────┘
               │
               ▼
┌─────────────────────────────────┐
│   FOUR SPECTRAL POINTS PLACE     │
│    ON FREQUENCY DOMAIN AND       │
│  3·PT DIGITAL  FILTERING PERFORM │
└─────────────────────────────────┘
               │
               ▼
┌─────────────────────────────────┐
│    MULTIPLY BY RANDOM            │
│    NOISE IF UNVOICED             │
└─────────────────────────────────┘
               │
               ▼
┌─────────────────────────────────┐
│      INVERSE FFT TO              │
│   GENERATE  WAVEFORM             │
│   FOR EACH PITCH PERIOD          │
└─────────────────────────────────┘
               │
               ▼
┌─────────────────────────────────┐
│     HIGHLY INTELLIGIBLE          │
│     SPEECH REPRODUCTION          │
│  AT 8KBIT/SEC D/A CONVERTER      │
└─────────────────────────────────┘
```

GENERAL
PURPOSE COMPUTER

$\boxed{\text{FIG. }7B}$

FIG . 8 .

50ms

FIG. 9 .

20Khz

FIG. 10 .

0132216

20Khz

FIG. 11.

60Khz

FIG. 12.

60Khz

FIG. 13.

10Khz

FIG. 14.

10Khz

FIG. 15.

FIG. 15a.

6Khz

0·5s

0132216

REAL TIME SPEECH SYNTHESIS BY
DECODING SPECTRAL PEAK INFORMATION.

SPECTRAL PEAKS INFORMATION FROM COMPUTER

↓

HP PLOTTER

↓

ROTATING DRUM

↓

OPTICAL READER USING RETICON CHIP

↓

A/D CONVERSION FOR EACH RESPECTIVE LINE
USING 8BITS/LINE FOR ADDRESSING FREQUENCY

↓

D/A CONVERSION USING LADDER –
NETWORK TO REGULATE CURRENT FLOW

↓

FOUR X2206 OSCILLATORS CHIPS USED TO CONVERT
POSITIONS OF LINES INTO FREQUENCIES

↓

OUTPUT FROM OSCILLATORS SUMMED

↓

OUTPUT MULTIPLIED BY TRIANGULAR FUNCTION

↓

INTELLIGIBLE SPEECH

↓

8BITS A/D CONVERSION TO
STORE DATA INTO COMPUTER

↓

SYNTHESIZED SPEECH REANALYSED
TO PRODUCE 3-DIMENSIONAL PLOT

FIG. 16

0132216

FIG. 18.

50ms

FIG. 19.

17ms

FIG. 20.

17ms

FIG. 21.

17ms

FIG. 22.

10Khz

FIG. 23.

10Khz

European Patent Office

EUROPEAN SEARCH REPORT

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl. ³)
X	ELECTRONICS AND COMMUNICATIONS IN JAPAN, vol. 62-A, no. 4, 1979, pages 10-17, Scripta Publishing Co., Washington, US; S. IMAI et al.: "Spectral envelope extraction by improved cepstral method" * paragraph 2.1 "Cepstral method" *	1,2,4, 6,12, 15,16	G 10 L 1/02
A	THE JOURNAL OF THE ACOUSTICAL SOCIETY OF JAPAN, vol. 32, no. 1, January 1976, pages 12-23, Tokyo, JP; T. MATSUOKA et al.: "Investigation on phonemic information of static properties of local peaks in the speech spectra" * abstract *	1,15	
			TECHNICAL FIELDS SEARCHED (Int. Cl. ³)
A	IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, vol. ASSP-22, no. 5, October 1974, pages 362-381, IEEE, New York, US; H.F. SILVERMAN et al.: "A parametrically controlled spectral analysis system for speech" * paragraph V: "Frequency-domain shaping" *	6-11	G 10 L 1/02
A	FR-A-2 337 393 (DIALOG SYSTEMS) * claim 1 *	1-3	
	--- -/-		

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	12-10-1984	ARMSPACH J.F.A.M.

EPO Form 1503. 03.82